

NONCOMMUTATIVE BENNETT AND ROSENTHAL INEQUALITIES

MARIUS JUNGE AND QIANG ZENG

ABSTRACT. In this paper we extend Bennett's and Bernstein's inequality to the noncommutative setting. In addition we provide an improved version of the noncommutative Rosenthal inequality, essentially due to Nagaev, Pinelis, and Pinelis, Utev for commutative random variables. We also present new best constants in Rosenthal's inequality. Applying these results to random Fourier projections, we recover and elaborate on fundamental results from compressed sensing, due to Candes, Romberg, and Tao.

0. INTRODUCTION

Rosenthal's inequality [34] was initially discovered to construct some new Banach spaces. However, Rosenthal's inequality gives a very nice bound for the p -norm of independent random variables, and has found many generalizations and applications. The martingale version of Rosenthal's inequality was discovered almost simultaneously by Burkholder [4]. Since then the order of the constant in these inequalities have been studied extensively, in particular by Johnson, Schechtman and Zinn [17]. The correct order in the martingale version has been established by Hitczenko [16], based on fundamental work of Kwapien and Woyczyński [23]. Nowadays, easy proofs of Rosenthal inequalities can be found with the help of Bennett and Bernstein's inequalities, see [3] and the references therein. We will extend Bennett's inequalities to the noncommutative setting. Let us recall that the classical Rosenthal inequality says that for independent mean 0 random variables we have

$$(0.1) \quad \left(\mathbb{E} \left| \sum_{k=1}^n f_k \right|^p \right)^{1/p} \leq c(p) \left(\left(\sum_{k=1}^n \mathbb{E} |f_k|^2 \right)^{1/2} + \left(\sum_{k=1}^n \mathbb{E} |f_k|^p \right)^{1/p} \right).$$

2010 *Mathematics Subject Classification.* Primary 46L53; Secondary 46L50, 60E15, 60F10, 94A12.

Key words and phrases. (Noncommutative) Bennett inequality, (noncommutative) Rosenthal inequality, (noncommutative) Bernstein inequality, noncommutative L_p spaces, compressed sensing, large deviation, Cramér's theorem.

The first author was Partially supported by NSF Grant DMS-0901457.

According to [17], the order of the best constant here is $c(p) = p/(1 + \log p)$. In this paper we separate the two terms and ask for

$$(0.2) \quad \left(\mathbb{E} \left| \sum_{k=1}^n f_k \right|^p \right)^{1/p} \leq A(p) \left(\sum_{k=1}^n \mathbb{E} |f_k|^2 \right)^{1/2} + B(p) \left(\sum_{k=1}^n \mathbb{E} |f_k|^p \right)^{1/p}.$$

The central limit theorem immediately implies $A(p) \geq c\sqrt{p}$ for every choice of $B(p)$. The problem (0.2) is by no means new. Nagaev and Pinelis [26] obtained a very precise bound on the tail behaviour of $S_n = \sum_{k=1}^n X_k$ which implies that $(A(p), B(p)) = C(\sqrt{p}, p)$ is possible (although it is not a trivial task to deduce this estimate from their original inequality). Pinelis and Utev showed that in some sense $A(p) = C\sqrt{p}$ and $B(p) = Cp$ is best possible. In section 3, we will revisit this problem and show that assuming $A(p) \leq Cp^m$ for some $m > 1/2$ we must have

$$B(p) \geq c \frac{p}{1 + \log p}.$$

This is exactly consistent with $(A(p), B(p)) = C(p/(1 + \log p), p/(1 + \log p))$. Moreover, we show that the worst case is obtained for independent random selectors $f_k = (\delta_k - \lambda)$ with expectation $\lambda > 0$.

We will prove a vast generalization of (0.2) in the noncommutative setting for conditionally independent random variables with $A(p) = c\sqrt{p}$ and $B(p) = Cp$. This improves the corresponding results from [22] of the form $A(p) = B(p) = Cp$. Our new results are motivated by applications in compressed sensing for random selectors with matrix valued coefficients. More precisely, we have to consider rank one operators

$$a_j = [\bar{x}_j(l)x_j(r)]_{1 \leq l, r \leq n}$$

such that $|x_k(j)| \leq D$. Then the aim is to estimate

$$(0.3) \quad \left\| \frac{1}{k} \sum_{j=1}^n \delta_j f a_j f - f \right\|_{B(\ell_2^n)} \leq ?$$

for independent selectors $\delta_j \in \{0, 1\}$ with $\mathbb{E}\delta_j = k/n$ and a projection f . As in the original paper [6] by Candes, Romberg, and Tao, a tempting approach is to use moment estimates, or equivalently estimates of the Schatten p -norm of these matrices. In fact, the improved Rosenthal inequality allows us to recover the famous estimates in [6].

Let us recall that the noncommutative L_p space associated with the trace on $B(\ell_2)$ is given by

$$\|x\|_p = [tr(|x|^p)]^{1/p} = \left(\sum_j s_j(x)^p \right)^{1/p}$$

where the singular number $s_j(x) = \lambda_j(|x|)$, i.e. the eigenvalues of the positive matrix $|x| = \sqrt{x^*x}$. Thus a good estimate of (0.3) can certainly be obtained from an estimate of the form

$$(0.4) \quad \left(\mathbb{E} \left\| \sum_j \delta_j f a_j f - k f \right\|_p^p \right)^{1/p} \leq C \sqrt{p} \left\| \sum_j \mathbb{E}(\delta_j^2) f a_j f^2 \right\|_{p/2}^{1/2} + C p \left(\sum_j \|f a_j f\|_p^p \right)^{1/p}.$$

Let us now describe the more general setup which allows us to prove results in noncommutative probability which includes all the statements above. Indeed, we assume that \mathcal{M} is a von Neumann algebra equipped with a normal faithful tracial state $\tau : \mathcal{M} \rightarrow \mathbb{C}$, i.e. $\tau(1) = 1$ and $\tau(xy) = \tau(yx)$. Then $L_p(\mathcal{M}, \tau)$ is the completion of \mathcal{M} with respect to $\|x\|_p = [\tau(|x|^p)]^{1/p}$. It is well-known (see for example [14, 31]) that $\|\cdot\|_p$ is a norm for $1 \leq p \leq \infty$. In particular, $\|\cdot\|_\infty = \|\cdot\|$. Here and in the following $\|\cdot\|$ will always denote the operator norm. Let $\mathcal{N} \subset \mathcal{M}$ be a von Neumann subalgebra. Then there exists a unique conditional expectation $E_{\mathcal{N}} : \mathcal{M} \rightarrow \mathcal{N}$ such that $E_{\mathcal{N}}(1) = 1$ and

$$E_{\mathcal{N}}(axb) = aE_{\mathcal{N}}(x)b \quad a, b \in \mathcal{N} \text{ and } x \in \mathcal{M}.$$

We say that two subalgebras $\mathcal{N} \subset A, B \subset \mathcal{M}$ are *independent* over \mathcal{N} if

$$E_{\mathcal{N}}(ab) = E_{\mathcal{N}}(a)E_{\mathcal{N}}(b) \quad a \in A, b \in B.$$

In particular, we say that $x, y \in \mathcal{M}$ are independent if the algebras they generate respectively are independent over \mathbb{C} . A sequence of subalgebra A_1, \dots, A_n are called successively independent over \mathcal{N} if A_{k+1} is independent of the algebra $\mathcal{M}(k)$ generated by A_1, \dots, A_k . Our noncommutative Bennett inequality reads as follows.

Theorem 0.1. *Let $\mathcal{N} \subset A_j \subset \mathcal{M}$ be successively independent over \mathcal{N} and $a_j \in A_j$ be self-adjoint such that*

$$\text{i) } E_{\mathcal{N}}(a_j) = 0; \text{ ii) } E_{\mathcal{N}}(a_j^2) \leq \sigma_j^2; \text{ iii) } \|a_j\| \leq M_j.$$

Then for $t \geq 0$,

$$\tau \left(1_{[t, \infty)} \left(\sum_{j=1}^n a_j \right) \right) \leq \exp \left(- \frac{\sum_{j=1}^n \sigma_j^2}{\sup_{j=1, \dots, n} M_j^2} \phi \left(\frac{t \sup_{j=1, \dots, n} M_j}{\sum_{j=1}^n \sigma_j^2} \right) \right)$$

where $\phi(x) = (1+x) \log(1+x) - x$.

Here we used $1_I(a) = \int_I dE_t$ for the spectral projection given by the spectral decomposition $a = \int t dE_t$. We should mention that the key new ingredient in this theorem is the Golden-Thompson inequality, which has already played a crucial role in Gross' paper

[15]. We invite the reader to rewrite the inequality for conditionally independent copies x_j with $\sigma = \sigma_j$, $M_j = M$. Note that in the commutative context

$$\tau(1_{[t,\infty)}(a)) = \text{Prob}(a \geq t).$$

In the future we will simply take this formula as a definition. Then our Bernstein inequality for noncommutative random variables reads as follows.

Corollary 0.2. *Under the same hypothesis of Theorem 0.1*

$$\text{Prob}\left(\sum_{j=1}^n a_j \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{j=1}^n \sigma_j^2 + \frac{2}{3}t \sup_{j=1,\dots,n} M_j}\right).$$

In the work of Ahlswede and Winter [1], and Gross [15], a similar, but different version of Bernstein's inequality was used. Indeed, in [1] the Bernstein inequality applies to independent random variables in a von Neumann algebra and gives information about the random spectrum. In our inequality we allow general randomness via independence, but obtain a slightly different conclusion. At any rate, from our version it is now rather standard to derive Rosenthal's inequality from Bernstein's inequality.

Corollary 0.3. *Let $2 \leq p < \infty$ and a_j satisfy the hypothesis of Theorem 0.1. Then*

$$\left\|\sum_{j=1}^n a_j\right\|_p \leq C \left(\left(p \sum_{j=1}^n \sigma_j^2\right)^{1/2} + p \sup_{j=1,\dots,n} M_j \right).$$

For unbounded operators and fixed p we can prove a similar inequality. Here we have to make a slightly stronger assumption. Let us recall that $(A_j)_{j=1}^n$ are fully independent over \mathcal{N} if for every subset $I \subset \{1, \dots, n\}$ the algebra $\mathcal{M}(I)$ generated by $\bigcup_{i \in I} A_i$ is independent from $\mathcal{M}(I^c)$ over \mathcal{N} .

Theorem 0.4. *Let (A_i) be fully independent over \mathcal{N} , $1 \leq p < \infty$, $x_i \in L_p(A_i)$ with $E_{\mathcal{N}}(x_i) = 0$. Then*

$$(0.5) \quad \left\|\sum_{j=1}^n x_j\right\|_p \leq C \max \left\{ \sqrt{p} \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^* + x_j^* x_j) \right)^{1/2} \right\|_p, p \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p} \right\}.$$

If moreover, $p \geq 2.5$ then

$$(0.6) \quad \left\|\sum_{j=1}^n x_j\right\|_p \leq C \max \left\{ \sqrt{p} \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^* + x_j^* x_j) \right)^{1/2} \right\|_p, p \|(x_j)\|_{L_p(\ell_\infty)} \right\}.$$

According to [29] and [19], the norm of (x_j) in $L_p(\ell_\infty)$ is given by $\inf\{\|a\|_{2p}\|b\|_{2p}\}$ such that

$$x_j = ay_jb \quad \text{with} \quad \|y_j\|_\infty \leq 1.$$

Clearly, the orders \sqrt{p} and p in the above theorem are optimal because they are already optimal in commutative probability. Note that in this version Theorem 0.4 improves on Corollary 0.3 for p large enough. The passage from first assertion to the second follows from an argument in [22]. In fact, Rosenthal inequalities in the noncommutative setting have been successively explored in [20, 21] and [22]. The martingale situation is completely settled due to the work of [32] which shows that for noncommutative martingales

$$\left\| \sum_j d_j \right\|_p \leq Cp \left(\left\| \left(\sum_k E_{k-1}(d_k^* d_k + d_k^* d_k) \right)^{1/2} \right\|_p + \left(\sum_k \|d_k\|_p^p \right)^{1/p} \right)$$

where (d_k) is a sequence of martingale differences given by $E_k(x) = E_{\mathcal{N}_k}(x)$ and $d_k = d_k(x) = E_k(x) - E_{k-1}(x)$ for a filtration $(\mathcal{N}_k) \subset \mathcal{M}$. As observed in [21] the constant Cp gives the correct order.

Let us return to the situation in compressed sensing. Here we obtain the following result.

Corollary 0.5. *Let $x_j \in \mathcal{N}$ be positive operator, τ a normalized trace such that*

$$\text{i) } \frac{1}{m} \sum_{j=1}^m x_j = 1; \text{ ii) } \|x_j\| \leq r.$$

Let δ_j be independent selectors such that $\mathbb{E}\delta_j = k/m$. Then for $p \geq 2.5$

$$(0.7) \quad \left(\mathbb{E} \left\| \frac{1}{k} \sum_{j=1}^m \delta_j x_j - 1 \right\|_{L_p(\tau)}^p \right)^{1/p} \leq C \max \left\{ \sqrt{\frac{pr}{k}}, \frac{pr}{k} \right\}.$$

Moreover, if tr is a trace on \mathcal{N} such that

$$\|x\|_{L_\infty(tr)} \leq \|x\|_{L_p(tr)},$$

and $r/k = \varepsilon^2$, then, for $t^2 \geq 2.5C^2e$ and $t \geq 2.5Ce\varepsilon$, we have

$$(0.8) \quad \text{Prob} \left(\left\| \frac{1}{k} \sum_{j=1}^m \delta_j x_j - 1 \right\|_{L_\infty(tr)} > t\varepsilon \right) \leq tr(1) \begin{cases} e^{-t^2/(2C^2e)} & \text{if } t\varepsilon \leq C \\ e^{-t/(2Ce\varepsilon)} & \text{if } t\varepsilon \geq C \end{cases}.$$

Here C is an absolute constant.

These results are closely related to the matrix Bernstein inequality from Tropp's paper [38] and operator Bernstein inequality from [15]. Their application to problem in compressed sensing will be explained in section 4. Section 1 provides the proof of the Bennett and Bernstein's inequalities. An application to large deviation inequalities and how non-commutative gaussian random variables may violate the classical equalities are discussed in section 2. The improved Rosenthal inequality is proved in section 3.

1. NONCOMMUTATIVE BENNETT'S INEQUALITY

Let us first recall some background. For a self-adjoint operator $a \in \mathcal{M}$, we have the spectral decomposition $a = \int t dE_t$, where E_t is the spectral measure of a . For any Borel set $A \subset \mathbb{R}$, we define $\mu(A) := \tau(E(A))$. Then μ is a scalar-valued spectral measure for a and $\mu(\mathbb{R}) = 1$. By the measurable functional calculus (see for example [11, Section IX.8]), there exists a $*$ -homomorphism $\pi : L^\infty(\mu) \rightarrow \mathcal{M}$ depending on a such that for all $f \in L^\infty(\mu)$, $\pi(f) = f(a)$, and

$$(1.1) \quad \tau(f(a)) = \int f(t) \mu(dt).$$

In particular, for $f = 1_{[t, \infty)}$, we have the exponential Chebyshev inequality

$$(1.2) \quad \tau(1_{[t, \infty)}(a)) = \text{Prob}(a \geq t) \leq e^{-t} \tau(e^a).$$

Our proof of Bennett's inequality relies on the following crucial result obtained in [36, Theorem 4], see also [2, 27] and the references therein.

Lemma 1.1 (Golden-Thompson inequality). *Suppose that a, b are self-adjoint operators, bounded above and that $a + b$ are essentially self-adjoint (i.e. the closure of $a + b$ is self-adjoint). Then*

$$\tau(e^{a+b}) \leq \tau(e^{a/2} e^b e^{a/2}).$$

Furthermore, if $\tau(e^a) < \infty$ or $\tau(e^b) < \infty$ then

$$(1.3) \quad \tau(e^{a+b}) \leq \tau(e^a e^b).$$

Note that if $a, b \in \mathcal{M}$ are self-adjoint, the hypotheses in Lemma 1.1 are automatically satisfied. Therefore we have (1.3). With the help of (1.2) and (1.3), we can prove the noncommutative Bennett inequality following the commutative case given in [3].

Proof of Theorem 0.1. (1.2) implies for $\lambda \geq 0$,

$$(1.4) \quad \text{Prob} \left(\sum_{i=1}^n a_i \geq t \right) \leq e^{-\lambda t} \tau(e^{\lambda \sum_{i=1}^n a_i}).$$

Since (a_i) are successively independent, we deduce from (1.3) that

$$(1.5) \quad \begin{aligned} \tau(e^{\lambda \sum_{i=1}^n a_i}) &\leq \tau(e^{\lambda \sum_{i=1}^{n-1} a_i} e^{\lambda a_n}) = \tau(E_{\mathcal{N}}(e^{\lambda \sum_{i=1}^{n-1} a_i} e^{\lambda a_n})) \\ &= \tau(E_{\mathcal{N}}(e^{\lambda \sum_{i=1}^{n-1} a_i}) E_{\mathcal{N}}(e^{\lambda a_n})). \end{aligned}$$

Expanding, we obtain

$$\begin{aligned} E_{\mathcal{N}}(e^{\lambda a_n}) &= E_{\mathcal{N}} \left(\sum_{k=0}^{\infty} \frac{(\lambda a_n)^k}{k!} \right) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} E_{\mathcal{N}}(a_n^k) \\ &= 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} E_{\mathcal{N}}(a_n^2 a_n^{k-2}) \leq 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} M_n^{k-2} \sigma_n^2 \\ &= 1 + \frac{\sigma_n^2}{M_n^2} (e^{\lambda M_n} - 1 - \lambda M_n) \leq \exp \left(\frac{\sigma_n^2}{M_n^2} (e^{\lambda M_n} - 1 - \lambda M_n) \right). \end{aligned}$$

Note that the function $f(x) := \exp(x^{-2}(e^{\lambda x} - 1 - \lambda x))$ is increasing for $x > 0$. It follows that

$$E_{\mathcal{N}}(e^{\lambda a_n}) \leq \exp \left(\frac{\sigma_n^2}{C^2} (e^{\lambda C} - 1 - \lambda C) \right)$$

where $C = \sup_{i=1, \dots, n} M_i$. Iterating $n-2$ times, we obtain

$$\tau(e^{\lambda \sum_{i=1}^n a_i}) \leq \exp \left(\frac{\sum_{i=1}^n \sigma_i^2}{C^2} (e^{\lambda C} - 1 - \lambda C) \right).$$

This yields

$$(1.6) \quad \text{Prob} \left(\sum_{i=1}^n a_i \geq t \right) \leq \exp \left(-\lambda t + \frac{\sum_{i=1}^n \sigma_i^2}{C^2} (e^{\lambda C} - 1 - \lambda C) \right).$$

By differentiating we find the minimizing value $\lambda = C^{-1} \log(1 + tC / (\sum_{i=1}^n \sigma_i^2))$. Then (1.6) yields the assertion. \square

It is known that Bernstein's inequality is a straightforward consequence of Bennett's inequality.

Proof of Corollary 0.2. Since $\phi(x) \geq x^2/(2 + 2x/3)$ for $x \geq 0$, the corollary follows by relaxing the bound in Bennett's inequality. \square

In the following we use Corollary 0.2 to prove Corollary 0.3. Let $a \in \mathcal{M}$ be positive. Recall that $\text{Prob}(a > t)$ is an analog of the classical distribution function of a . In particular we may use it to compute the L_p norm of a .

Lemma 1.2. *Assume that $p > 0$ and that $a \in \mathcal{M}$ is positive. Then*

$$\|a\|_p^p = p \int_0^\infty t^{p-1} \text{Prob}(a > t) dt.$$

Proof. According to Fubini theorem and (1.1),

$$\begin{aligned} \|a\|_p^p &= \tau(a^p) = \int_{\mathbb{R}} s^p \mu(ds) \\ &= \int_0^\infty \mu(t, \infty) p t^{p-1} dt = p \int_0^\infty t^{p-1} \text{Prob}(a > t) dt. \end{aligned} \quad \square$$

Recall that the Gamma function is defined as $\Gamma(p) = \int_0^\infty e^{-r} r^{p-1} dr$ and the incomplete Gamma function is defined as $\Gamma(\alpha, p) = \int_p^\infty e^{-t} t^{\alpha-1} dt$. We need an elementary estimate for $\Gamma(\alpha, p)$. Note that for $t \geq p \geq 2(\alpha - 1)$, we have

$$(e^{-t} t^{\alpha-1})' = -e^{-t} t^{\alpha-1} \left(1 - \frac{\alpha-1}{t}\right) \leq -\frac{1}{2} e^{-t} t^{\alpha-1}.$$

This gives the following lemma.

Lemma 1.3. *If $p \geq 2\alpha - 2$, then $\Gamma(\alpha, p) \leq 2e^{-p} p^{\alpha-1}$.*

Proof of Corollary 0.3. First note that symmetry and Corollary 0.2 imply

$$\text{Prob} \left(\left| \sum_{i=1}^n a_i \right| \geq t \right) \leq 2 \exp \left(-\frac{t^2}{2 \sum_{i=1}^n \sigma_i^2 + \frac{2}{3} t \sup_{1 \leq i \leq n} M_i} \right).$$

Put $S = \sum_{i=1}^n \sigma_i^2$ and $R = \sup_{i=1, \dots, n} M_i$. By Lemma 1.2, we have

$$\begin{aligned} \left\| \sum_{i=1}^n a_i \right\|_p^p &\leq 2p \int_0^\infty \exp \left(-\frac{t^2}{2S + \frac{2}{3} t R} \right) t^{p-1} dt \\ &= 2p \int_0^{\frac{3S}{R}} \exp \left(-\frac{t^2}{2S + \frac{2}{3} t R} \right) t^{p-1} dt + 2p \int_{\frac{3S}{R}}^\infty \exp \left(-\frac{t^2}{2S + \frac{2}{3} t R} \right) t^{p-1} dt \\ &= 2p(I + II) \end{aligned}$$

where

$$I = \int_0^{\frac{3S}{R}} \exp \left(-\frac{t^2}{2S + \frac{2}{3} t R} \right) t^{p-1} dt \quad \text{and} \quad II = \int_{\frac{3S}{R}}^\infty \exp \left(-\frac{t^2}{2S + \frac{2}{3} t R} \right) t^{p-1} dt.$$

We first estimate I . Since $t \leq 3S/R$, we have

$$I \leq \int_0^{\frac{3S}{R}} e^{-t^2/(4S)} t^{p-1} dt = 2^{p-1} S^{p/2} \int_0^{\frac{9S}{4R^2}} e^{-r} r^{p/2-1} dr.$$

For $9S/(4R^2) \leq p$, we have $I \leq 2^{p-1} S^{p/2} \int_0^p e^{-r} r^{p/2-1} dr \leq 2^p S^{p/2} p^{p/2-1}$. For $9S/(4R^2) > p$, we have

$$I \leq 2^{p-1} S^{p/2} \left(\int_0^p e^{-r} r^{p/2-1} dr + \int_p^{\frac{9S}{4R^2}} e^{-r} r^{p/2-1} dr \right) \leq 2^p S^{p/2} p^{p/2-1} + I_2$$

where $I_2 = 2^{p-1} S^{p/2} \int_p^\infty e^{-r} r^{p/2-1} dr$ and by Lemma 1.3, $I_2 \leq 2^p S^{p/2} p^{p/2-1} e^{-p}$. Hence, we obtain

$$I \leq 2^{p+1} S^{p/2} p^{p/2-1}.$$

To estimate II , since $2S < 2tR/3$, we have

$$\begin{aligned} II &\leq \int_{\frac{3S}{R}}^\infty e^{-3t/(4R)} t^{p-1} dt = \left(\frac{4}{3} R \right)^p \int_{\frac{9S}{4R^2}}^\infty e^{-r} r^{p-1} dr \\ &\leq \left(\frac{4}{3} R \right)^p \Gamma(p) \leq \left(\frac{4}{3} R p \right)^p. \end{aligned}$$

Combining all the inequalities together, we find $\|\sum_{i=1}^n a_i\|_p^p \leq 2^{p+2} S^{p/2} p^{p/2} + 2(4R/3)^p p^{p+1}$. Hence, we obtain

$$\left\| \sum_{i=1}^n a_i \right\|_p \leq 4\sqrt{Sp} + \frac{4\sqrt{2}}{3} e^{1/e} R p \leq 4(\sqrt{Sp} + R p). \quad \square$$

We remark that the constant in the above inequality is explicit and quite small, which may be good for numerical purpose.

2. LARGE DEVIATION PRINCIPLE

Bennett's inequality is a large deviation type inequality giving an upper bound for the tail probability. In the commutative setting lower bounds have been analyzed intensively in Large Deviation Theory. Despite the fact that our arguments in the previous section are almost commutative, lower bounds for noncommutative random variables are very different. Let us start with Cramér's Theorem. We consider a sequence of fully independent and identically distributed (IID) noncommutative random variables $(a_i)_{i \in I}$.

Let $\Lambda(\lambda) = \log \tau(e^{\lambda a_1})$. Following [12] we define the Fenchel-Legendre transform of $\Lambda(\lambda)$ for $x \in \mathbb{R}$

$$(2.1) \quad \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} [\lambda x - \Lambda(\lambda)].$$

If (a_i) is a commutative IID sequence, then Cramér's Theorem [12, Theorem 2.2.3] says that (a_i) satisfies the Large Deviation Principle (LDP) with rate function Λ^* , which implies [12, Corollary 2.2.19]

$$(2.2) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log \text{Prob} \left(\sum_{i=1}^n a_i \geq nt \right) = -\inf_{s \geq t} \Lambda^*(s).$$

The upper bound remains valid in the noncommutative setting.

Proposition 2.1. *Let $(a_i)_{i \geq 1}$ be an IID sequence in (\mathcal{M}, τ) such that $\tau(a_i) = 0$ for all $i \geq 1$. Then for any $t > 0$,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \text{Prob} \left(\sum_{i=1}^n a_i \geq nt \right) \leq -\inf_{s \geq t} \Lambda^*(s).$$

Proof. Thanks to the Golden-Thompson inequality, we can follow the proof in the commutative case in [12]. Using (1.4) and (1.5), we obtain

$$\text{Prob} \left(\sum_{i=1}^n a_i \geq nt \right) \leq e^{-\lambda nt} \prod_{i=1}^n \tau(e^{\lambda a_i}) = e^{-n(\lambda t - \Lambda(\lambda))}.$$

This implies

$$\frac{1}{n} \log \text{Prob} \left(\sum_{i=1}^n a_i \geq nt \right) \leq -\Lambda^*(t) \leq -\inf_{s \geq t} \Lambda^*(s). \quad \square$$

Remark 2.2. Although we assumed a_i 's are in (\mathcal{M}, τ) , using truncation and approximation, we can also prove the previous proposition for symmetric gaussians. To be more precise, for independent symmetric gaussian random variables a and b , let $a_N = a1_{\{|a| < N\}}$ and $b_N = b1_{\{|b| < N\}}$. Then Monotone Convergence Theorem implies that $\tau(e^{a_N}) \rightarrow \tau(e^a)$, $\tau(e^{b_N}) \rightarrow \tau(e^b)$. Since the symmetric gaussian random variable is in $\bigcap_{p \geq 1} L_p(\mathcal{M}, \tau)$, the triangle inequality implies $\tau((a_N + b_N)^p) \rightarrow \tau((a + b)^p)$. By symmetry, we have

$$\tau(e^{a_N + b_N}) \rightarrow \tau(e^{a+b}).$$

In the following we give two examples which violate the LDP for noncommutative random variables.

Example 2.3 (Noncommutative semicircular law [39]). Recall that the *semicircular law* centered at $a \in \mathbb{R}$ and of radius $r > 0$ is the distribution $\gamma_{a,r} : \mathbb{C}[X] \rightarrow \mathbb{C}$ defined by

$$\gamma_{a,r}(P) = \frac{2}{\pi r^2} \int_{a-r}^{a+r} P(t) \sqrt{r^2 - (t-a)^2} dt.$$

Here $\mathbb{C}[X]$ is the algebra of complex polynomials in one variable.

Let us recall that copies of semicircular random variables can be constructed on the full Fock space. Let H be a real Hilbert space and $H_{\mathbb{C}}$ its complexification. Let $\mathcal{T}(H_{\mathbb{C}})$ be the full Fock space on $H_{\mathbb{C}}$. For any $h \in H$ define $s(h) = l(h) + l(h)^*$, with $l(h)$ the left creation operator. Let $\Phi(H)$ be the von Neumann algebra generated by $\{s(h) | h \in H\}$. Let τ_H denote the vector state on $\Phi(H)$ given by the vacuum vector, $1 \in \mathcal{T}(H_{\mathbb{C}})$. Then for any orthonormal system $(h_i)_{i \in I} \subset H$ the family of random variables $(s_i) = s(h_i)$ is free (thus fully independent) in $(\Phi(H), \tau_H)$ and the distribution of $s(h)$ is the semicircular law $\gamma_{0,2}$.

By rotation invariance of the free functor we deduce from [39, Section 3.4] that

$$(2.3) \quad \hat{s}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n s_i \sim \gamma_{0,2},$$

which means that the distribution of \hat{s}_n is $\gamma_{0,2}$. Since $\gamma_{0,2}$ is supported in $[-2, 2]$, for any $t > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \text{Prob} \left(\sum_{i=1}^n s_i \geq nt \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \text{Prob}(\hat{s}_n \geq \sqrt{n}t) = -\infty.$$

On the other hand, by the integral representation of the modified Bessel function I_1 [37, (9.46)], the moment generating function of $\gamma_{0,2}$ is given by

$$M(\lambda) = \frac{1}{2\pi} \int_{-2}^2 e^{\lambda t} \sqrt{4 - t^2} dt = \frac{I_1(2\lambda)}{\lambda}.$$

Using the series representation of I_1 [37, (9.28)], we have for $\lambda > 0$,

$$M(\lambda) = \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{(n+1)!n!} \geq \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{2(2n)!} = \frac{e^{\lambda} + e^{-\lambda}}{4} \geq \frac{1}{4} e^{\lambda}.$$

We find $\Lambda(\lambda) = \log M(\lambda) \geq \lambda - \log 4$. Since $\tau(a_1) = 0$, by [12, Lemma 2.2.5], for $x \geq 0$,

$$\Lambda^*(x) = \sup_{\lambda \geq 0} [\lambda x - \Lambda(\lambda)].$$

Therefore,

$$\Lambda^*(1) = \sup_{\lambda \geq 0} (\lambda - \Lambda(\lambda)) \leq \log 4 < \infty,$$

which shows that the sequence (s_i) violates the LDP lower bound in (2.2). We have proved the following result.

Proposition 2.4. *The semicircular sequence $(s_n)_{n \in \mathbb{N}}$ does not satisfy LDP (2.2).*

The counterexample works in free probability because s_1 is bounded. In order to motivate the next example, we first clarify the relationship between the logarithmic moment generating function Λ and the rate function I of the LDP.

Remark 2.5. Suppose that an IID sequence (a_n) satisfies the LDP with rate function $I(x)$ and that $\Lambda(\lambda)$ is well-defined and lower semicontinuous. Then the Fenchel-Legendre transform of $I(x)$ coincides with $\Lambda(\lambda)$, i.e.

$$I^*(\lambda) \leq \Lambda(\lambda).$$

Indeed, by Hölder's inequality $\Lambda(\lambda)$ is convex. Then Cramér's Theorem and the duality lemma [12, Lemma 4.5.8] yield the assertion. In particular, if (a_n) satisfies the LDP with rate function I and Λ is lower semicontinuous, then $I(x) = x^2/2$ implies $\Lambda(\lambda) = I^*(\lambda) = \lambda^2/2$, i.e. the sequence (a_n) follows standard normal distribution. We have seen that in classical probability *the distribution of an IID sequence can be recovered from the rate function* given by the LDP provided that Λ is well-defined and lower semicontinuous. The next example will show that this is no longer true in the noncommutative setting.

Example 2.6 (Gaussian family). For $\theta \in (0, 1)$, given a noncommutative standard gaussian random variable g_0 ($\tau(g_0) = 0$ and $\tau(g_0^2) = 1$) and a noncommutative semicircular random variable $g_1 \sim \gamma_{0,2}$, one can construct (see [10]) a new sequence of IID noncommutative random variables $(\xi_i)_{i \geq 1}$ with the same distribution as g_θ such that for all $k \in \mathbb{N}$

$$\tau(g_\theta^k) = (1 - \theta)\tau(g_0^k) + \theta\tau(g_1^k).$$

This implies by approximation (see [18])

$$\tau(f(g_\theta)) = (1 - \theta)\tau(f(g_0)) + \theta\tau(f(g_1))$$

for all measurable function f . In particular, for all $x \in \mathbb{R}$,

$$(2.4) \quad \text{Prob}(g_\theta \geq x) = (1 - \theta)\text{Prob}(g_0 \geq x) + \theta\text{Prob}(g_1 \geq x)$$

and for all $\lambda \in \mathbb{R}$,

$$(2.5) \quad \tau(e^{\lambda g_\theta}) = (1 - \theta)\tau(e^{\lambda g_0}) + \theta\tau(e^{\lambda g_1}).$$

By (2.4) and the invariance property (2.3), we have

$$\text{Prob}\left(\sum_{i=1}^n \xi_i \geq nx\right) = \text{Prob}(g_\theta \geq \sqrt{n}x) = (1 - \theta)\text{Prob}(g_0 \geq \sqrt{n}x) + \theta\text{Prob}(g_1 \geq \sqrt{n}x).$$

Therefore, we obtain the large deviation identity

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \text{Prob} \left(\sum_{i=1}^n \xi_i \geq nx \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \text{Prob}(g_0 \geq \sqrt{n}x) = -\frac{x^2}{2}.$$

On the other hand, if we put $\Lambda_\theta(\lambda) = \log \tau(e^{\lambda g_\theta})$ and let ν denote the probability measure of semicircular law $\gamma_{0,2}$, then (2.5) implies

$$\begin{aligned} \Lambda_\theta(\lambda) &= \log \left((1-\theta)e^{\lambda^2/2} + \theta \int_{-2}^2 e^{\lambda t} \nu(dt) \right) \\ &\leq \log(1-\theta) + \frac{\lambda^2}{2} + \log \left(1 + \frac{\theta}{1-\theta} e^{2\lambda - \lambda^2/2} \right) \end{aligned}$$

and

$$\Lambda_\theta(\lambda) \geq \log(1-\theta) + \frac{\lambda^2}{2} + \log \left(1 + \frac{\theta}{1-\theta} e^{-2\lambda - \lambda^2/2} \right).$$

As $\lambda \rightarrow \infty$, we obtain

$$(2.6) \quad \left| \Lambda_\theta - \frac{\lambda^2}{2} - \log(1-\theta) \right| \leq \frac{\theta}{1-\theta} e^{-c\lambda^2}$$

where c is an absolute constant. But $\Lambda_0(\lambda) = \lambda^2/2$ is the logarithmic moment generating function of standard gaussian distribution. Let us record this as follows.

Proposition 2.7. *The sequence $(\xi_n)_{n \geq 1}$ satisfies the LDP (2.2) with rate function $I(x) = x^2/2$. However, the logarithmic moment generating function of ξ_1 differs from $\Lambda_0(\lambda) = \lambda^2/2$ as shown in (2.6).*

We have seen that the law of (ξ_n) can not be recovered from the LDP rate function. In view of Remark 2.5 and Proposition 2.7 we understand that LDP is a measure of commutativity.

3. IMPROVED NONCOMMUTATIVE ROSENTHAL'S INEQUALITY

We prove the improved noncommutative Rosenthal inequality and show that the coefficients can not be improved in this section. In order to prove Theorem 0.4, we will follow and refine the standard iteration procedure given in [22].

Proof of Theorem 0.4. Instead of proving (0.5) directly, we prove the following equivalent inequality

$$(3.1) \quad \left\| \sum_{j=1}^n x_j \right\|_p \leq D_p \max \left\{ \sqrt{p} \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j^* x_j) \right)^{1/2} \right\|_p, \right. \\ \left. \sqrt{p} \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^*) \right)^{1/2} \right\|_p, p \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p} \right\}$$

and we assume at the moment that D_p is the best constant which may depend on the range of p . By [22, Theorem 2.1], (3.1) is true for $1 \leq p \leq 4$. This is the starting point of our iteration argument. Assume $p > 2$. We only need to show “ $p \Rightarrow 2p$ ”. Let $x_i \in L_{2p}(\mathcal{M}, \tau)$. Write the conditional expectation operator $E = E_{\mathcal{N}}$ in the following proof. Put

$$A := \sqrt{2p} \left\| \left(\sum_{i=1}^n E(x_i^* x_i) \right)^{1/2} \right\|_{2p} \quad \text{and} \quad B := 2p \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/(2p)}.$$

Using [20, Lemma 1.2] and the noncommutative Khintchine inequality in [29] with the right order of best constant, we have

$$\left\| \sum_{i=1}^n x_i \right\|_{2p} \leq 2\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_{2p} \leq c\sqrt{p} \max \left\{ \left\| \sum_{i=1}^n x_i^* x_i \right\|_p^{1/2}, \left\| \sum_{i=1}^n x_i x_i^* \right\|_p^{1/2} \right\},$$

where (ε_i) is a sequence of Rademacher random variables and \mathbb{E} denotes the corresponding expectation. Let $y_i = x_i^* x_i - E(x_i^* x_i)$. Then

$$\left\| \sum_{i=1}^n x_i^* x_i \right\|_p \leq 2 \max \left\{ \left\| \sum_{i=1}^n y_i \right\|_p, \left\| \sum_{i=1}^n E(x_i^* x_i) \right\|_p \right\}.$$

Applying the induction hypothesis, we obtain

$$\left\| \sum_{i=1}^n y_i \right\|_p \leq D_p \max \left\{ \sqrt{p} \left\| \left(\sum_{i=1}^n E(y_i^2) \right)^{1/2} \right\|_p, p \left(\sum_{i=1}^n \|y_i\|_p^p \right)^{1/p} \right\}.$$

Note that

$$E(y_i^2) = E(|x_i|^4) - E(|x_i|^2)^2 \leq E(|x_i|^4).$$

By [20, Lemma 5.2], we obtain

$$\begin{aligned} \left\| \sum_{i=1}^n E(|x_i|^4) \right\|_{p/2} &\leq \left\| \sum_{i=1}^n E(|x_i|^2) \right\|_p^{(p-2)/(p-1)} \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/(p-1)} \\ &= (A^2/2p)^{(p-2)/(p-1)} (B/2p)^{2p/(p-1)} \\ &= A^{(2p-4)/(p-1)} B^{2p/(p-1)} (2p)^{-(3p-2)/(p-1)}. \end{aligned}$$

On the other hand, since E is a contraction on $L_p(\mathcal{M}, \tau)$, we have

$$\begin{aligned} \left(\sum_{i=1}^n \|y_i\|_p^p \right)^{1/p} &= \left(\sum_{i=1}^n \|x_i^* x_i - E(x_i^* x_i)\|_p^p \right)^{1/p} \\ &\leq 2 \left(\sum_{i=1}^n \|x_i^* x_i\|_p^p \right)^{1/p} = 2 \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/p} = \frac{B^2}{2p^2}. \end{aligned}$$

This gives

$$\begin{aligned} \left\| \sum_{i=1}^n y_i \right\|_p &\leq D_p \max\{\sqrt{p} A^{(p-2)/(p-1)} B^{p/(p-1)} (2p)^{-(3p-2)/(2p-2)}, p B^2/(2p^2)\} \\ &= D_p \max\{2^{-1/2} A^{(p-2)/(p-1)} B^{p/(p-1)} (2p)^{-1-1/(2p-2)}, B^2/(2p)\}. \end{aligned}$$

Hence, we find

$$(3.2) \quad \left\| \sum_{i=1}^n x_i^* x_i \right\|_p \leq 2 \max\{2^{-1/2} D_p A^{(p-2)/(p-1)} B^{p/(p-1)} (2p)^{-1-1/(2p-2)}, D_p B^2/(2p), A^2/(2p)\}.$$

Young's inequality for products implies

$$(3.3) \quad A^{(p-2)/(2p-2)} B^{p/(2p-2)} \leq A + B \leq 2 \max\{A, B\}.$$

(3.2) and (3.3) yield

$$\begin{aligned} \sqrt{p} \left\| \sum_{i=1}^n x_i^* x_i \right\|_p^{1/2} &\leq \max\{2^{1/4} \sqrt{D_p} \max\{A, B\}, \sqrt{D_p} B, A\} \leq 2^{1/4} \sqrt{D_p} \max\{A, B\} \\ &= 2^{1/4} \sqrt{D_p} \max\left\{ \sqrt{2p} \left\| \left(\sum_{i=1}^n E(x_i^* x_i) \right)^{1/2} \right\|_{2p}, 2p \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/(2p)} \right\}. \end{aligned}$$

Here we assumed $D_p \geq 2^{-1/2}$ without loss of generality. Applying the same argument to $x_i x_i^*$, we obtain

$$\sqrt{p} \left\| \sum_{i=1}^n x_i x_i^* \right\|_p^{1/2} \leq 2^{1/4} \sqrt{D_p} \max \left\{ \sqrt{2p} \left\| \left(\sum_{i=1}^n E(x_i x_i^*) \right)^{1/2} \right\|_{2p}, 2p \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/(2p)} \right\}.$$

Hence, (3.1) is true for $2p$ with constant $2^{1/4} c \sqrt{D_p}$. It follows that

$$D_{2p} \leq 2^{1/4} c \sqrt{D_p}$$

and thus $D_p \leq \sqrt{2} c^2$ which is independent of p . Therefore, the iteration argument is done and we have proved the first assertion. As mentioned in the introduction of this paper, the interpolation argument from [22, section 4] shows that the first assertion improves to the second assertion with a singularity as p tends to 2. Thus for $p \geq 2.5$ the assertion holds with an absolute constant. \square

Remark 3.1. The improved Rosenthal inequality allows us to extend Lust-Piquard's non-commutative Khintchine inequality [24, 25] in a twisted setting. We refer to [9] for unexplained notion on the gaussian measure space construction. The starting point is a discrete group acting on a real Hilbert space H . This means we fix an isometry $b : H \rightarrow L_2(\Omega, \Sigma, \mu)$ such that b is linear and $b(h)$ is a centered gaussian random variable with variance $\|h\|^2$. For example for $H = L_2(0, \infty)$ and $B_t = b(1_{[0,t]})$ we recover a well-known method to construct Brownian motion. We may assume that Σ is the minimal sigma algebra generated by the random variables $b(H)$. Then the action of G extends to a family of measure preserving automorphism $\alpha : G \rightarrow \text{Aut}(L_\infty(\Omega, \Sigma, \mu))$ such that

$$\alpha_g(b(h)) = b(g.h).$$

This allows us to form the crossed product $M = L_\infty(\Sigma) \rtimes G$. The crossed product is spanned by random variables of the form

$$x = \sum_g f_g \lambda(g).$$

Here $\lambda(g)$ refers to the regular representation of group. The algebraic structure is determined by $\lambda(g)f\lambda(g^{-1}) = \alpha_g(f)$. The twisted gaussian random variables are of the form

$$B = \sum_g b(h_g) \lambda(g) \quad , \quad h_g \in H.$$

In order to formulate the Khintchine inequality we have to recall that there exists trace preserving conditional expectation $E : M \rightarrow L(G)$. Here $L(G)$ is the von Neumann

subalgebra generated by the image $\lambda(G)$ and the trace is given by

$$\tau \left(\sum_g f_g \lambda(g) \right) = \int f_1 d\mu.$$

Then we can deduce from Theorem 0.4 that for $p \geq 2$

$$(3.4) \quad \|B\|_p \leq c\sqrt{p} \|E(B^*B + BB^*)^{1/2}\|_p.$$

Moreover, the span of the generalized gaussian random variables is complemented and the inequality remains true with additional vector valued coefficients. This is a key fact in proving noncommutative Riesz transforms. To illustrate (3.4) let us assume that the action is trivial. Let (e_k) be a basis and

$$B = \sum_{k,g} a(k,g) b(e_k) \otimes \lambda(g) = \sum_k b(e_k) \otimes a_k.$$

Then we find

$$E(BB^*) = \sum_k a_k a_k^* \quad , \quad E(B^*B) = \sum_k a_k^* a_k.$$

Thus the right hand side gives exactly the square function we expect for gaussian variables. However, with non-trivial additional group action BB^* and B^*B look quite different and the group action interferes significantly.

Using (0.6), we can prove Corollary 0.5 which will play a central role in the application to compressed sensing in the next section.

Proof of Corollary 0.5. By Jensen's inequality, we have

$$\begin{aligned} & \left(\mathbb{E}_\delta \left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_p(\mathcal{N}, \tau)}^p \right)^{1/p} = \left(\mathbb{E}_\delta \left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - \frac{1}{k} \mathbb{E}_{\delta'} \left(\sum_{i=1}^m \delta'_i x_i \right) \right\|_{L_p(\mathcal{N}, \tau)}^p \right)^{1/p} \\ & \leq \left(\mathbb{E}_\delta \left(\mathbb{E}_{\delta'} \left\| \frac{1}{k} \sum_{i=1}^m (\delta_i - \delta'_i) x_i \right\|_{L_p(\mathcal{N}, \tau)}^p \right)^p \right)^{1/p} \leq \left(\mathbb{E}_{\delta, \delta'} \left\| \frac{1}{k} \sum_{i=1}^m (\delta_i - \delta'_i) x_i \right\|_{L_p(\mathcal{N}, \tau)}^p \right)^{1/p} \end{aligned}$$

where (δ'_i) is a sequence of independent selectors with the same distribution as δ_i 's. In order to apply Theorem 0.4, it is crucial to choose appropriate probability space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space generated by (δ_i, δ'_i) . We consider the noncommutative probability space as the algebra $\mathcal{M} = L_\infty(\mathbb{P}) \otimes \mathcal{N}$. Then we have a normalized trace $\tilde{\tau} = \mathbb{E} \otimes \tau$ on \mathcal{M} . We identify \mathbb{E} as the conditional expectation $\mathbb{E} : \mathcal{M} \rightarrow \mathcal{N}$. Clearly,

$((\delta_i - \delta'_i)x_i)_{i=1}^n$ are fully independent over \mathcal{N} . Note that

$$\mathbb{E}(\delta_i - \delta'_i)^2 = \frac{2k}{m} \left(1 - \frac{k}{m}\right) \leq \frac{2k}{m} \quad \text{and} \quad \sup_{i=1, \dots, m} |\delta_i - \delta'_i| \leq 1.$$

Since x_i is positive, $x_i^* x_i = x_i^2$. Using (0.6), we obtain

$$\begin{aligned} & \left(\mathbb{E} \left\| \sum_{i=1}^m (\delta_i - \delta'_i) x_i \right\|_{L_p(\mathcal{N}, \tau)}^p \right)^{1/p} = \left\| \sum_{i=1}^m (\delta_i - \delta'_i) x_i \right\|_{L_p(\mathcal{M}, \tilde{\tau})} \\ & \leq C \max \left\{ \sqrt{p} \left\| \sum_{i=1}^m \mathbb{E}((\delta_i - \delta'_i)^2 x_i^2) \right\|_{L_{p/2}(\mathcal{N}, \tau)}^{1/2}, p \left\| \sup_{i=1, \dots, m} |\delta_i - \delta'_i| x_i \right\|_{L_p(\mathcal{M}, \tilde{\tau})} \right\}. \end{aligned}$$

Since $\tau(1) = 1$ and $x_i \leq r$, we obtain $\|\delta_i - \delta'_i\|_{L_p(\mathcal{M}, \tilde{\tau}; l_\infty)} \leq r$, and

$$\left\| \sum_{i=1}^m \mathbb{E}(\delta_i - \delta'_i)^2 x_i^2 \right\|_{L_{p/2}(\mathcal{N}, \tau)} \leq 2kr \left\| \frac{1}{m} \sum_{i=1}^m x_i \right\|_{L_{p/2}(\mathcal{N}, \tau)} = 2kr.$$

Therefore, we find

$$\left(\mathbb{E} \left\| \frac{1}{k} \sum_{i=1}^m (\delta_i - \delta'_i) x_i \right\|_{L_p(\mathcal{N}, \tau)}^p \right)^{1/p} \leq C \max \left\{ \sqrt{\frac{2pr}{k}}, \frac{pr}{k} \right\}.$$

We have completed the proof of (0.7) with constant $\sqrt{2}C$. For the “moreover” part, we use the additional norm assumption and obtain

$$\left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_\infty(tr)} \leq \left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_p(tr)}.$$

Then by Chebyshev’s inequality and (0.7) for trace $\tau(x) = tr(x)/tr(1)$, we have

$$\begin{aligned} \mathbb{P} \left(\left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_\infty(tr)} \geq t\varepsilon \right) & \leq (t\varepsilon)^{-p} \mathbb{E} \left\| \frac{1}{k} \sum_{i=1}^m (\delta_i - \delta'_i) x_i \right\|_{L_p(tr)}^p \\ & \leq tr(1) \max \left\{ \sqrt{\frac{C^2 pr}{kt^2 \varepsilon^2}}, \frac{Cpr}{kt\varepsilon} \right\}^p. \end{aligned}$$

Let us first assume $t\varepsilon \leq C$. Optimize the first term in p and find $p = t^2 \varepsilon^2 k / (C^2 r e)$. Recall that $k = r\varepsilon^{-2}$. Then the first term becomes $e^{-t^2/(2C^2 e)}$. Using $t\varepsilon \leq C$, this choice of p gives an upper bound of $e^{-t^2/(C^2 e)}$ for the second term. Now assume $t\varepsilon \geq C$. The optimal choice for the second term is obtained for $p = kt\varepsilon / (C r e)$. Then the second term becomes $e^{-t/(C e \varepsilon)}$

and, thanks to $t\varepsilon \geq C$, the first term is less than $e^{-t/(2Ce\varepsilon)}$. The additional assumption on t guarantees that $p \geq 2.5$ in both cases. Therefore,

$$\mathbb{P} \left(\left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_\infty(tr)} \geq t\varepsilon \right) \leq tr(1) \begin{cases} e^{-t^2/(2C^2e)} & \text{if } t\varepsilon \leq C, \\ e^{-t/(2Ce\varepsilon)} & \text{if } t\varepsilon \geq C. \end{cases}$$

The constant C is the same as the constant in the first assertion. \square

Remark 3.2. In this context it is useful to compare our different generalizations of Rosenthal's inequality. We observe that with Corollary 0.3, we can only obtain

$$\left(\mathbb{E} \left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_p(\tau)}^p \right)^{1/p} \leq C \left(\sqrt{\frac{pr^2}{k}} + \frac{pr}{k} \right),$$

and with inequality (0.5) we obtain

$$\left(\mathbb{E} \left\| \frac{1}{k} \sum_{i=1}^m \delta_i x_i - 1 \right\|_{L_p(\tau)}^p \right)^{1/p} \leq C \left(\sqrt{\frac{pr}{k}} + \frac{pr}{k^{1-1/p}} \right).$$

Both estimates are worse than inequality (0.7).

The following two examples are meant to justify the optimality of \sqrt{p} and p . We refer the reader to [28] for a more detailed discussion on this topic in the framework of classical probability. We will use the standard notations for comparing orders of functions as $p \rightarrow \infty$. Recall that $f(p) = O(g(p))$ if there exists a constant C such that $f(p) \leq Cg(p)$ asymptotically, $f(p) = \Omega(g(p))$ if there exists a constant c such that $f(p) \geq cg(p)$ asymptotically, $f(p) = \Theta(g(p))$ if there exist constants c and C such that $cg(p) \leq f(p) \leq Cg(p)$ asymptotically, and $f(p) \sim g(p)$ if $\lim_{p \rightarrow \infty} f(p)/g(p) = 1$.

Example 3.3 (The optimality of \sqrt{p} in Theorem 0.4). Let us assume that

$$(3.5) \quad \left\| \sum_{i=1}^n x_i \right\|_p \leq A(p) \left(\sum_{i=1}^n \|x_i\|^2 \right)^{1/2} + B(p) \left(\sum_{i=1}^n \|x_i\|^p \right)^{1/p}$$

for some functions $A(p)$ and $B(p)$. We use $x_i = g_i$. Here (g_i) is a sequence of IID normal random variables with mean 0 and variance 1. We know $\mathbb{E}|g_1|^p = \frac{2^{p/2}}{\sqrt{\pi}} \Gamma(\frac{p+1}{2})$. By Stirling's formula, we obtain for large p ,

$$\|g_1\|_p \sim \sqrt{\frac{p}{e}}.$$

This yields that there exist absolute constants c and C such that $c\sqrt{p} \leq \|g_1\| \leq C\sqrt{p}$ for all $p \geq 2$. Hence, we obtain

$$c\sqrt{p} \leq \|g_1\|_p = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \right\|_p \leq A(p) + CB(p)\sqrt{p}n^{\frac{1}{p}-\frac{1}{2}}.$$

Sending $n \rightarrow \infty$, we have

$$A(p) \geq c\sqrt{p} \quad \text{for } p > 2.$$

This shows that one can not reduce the order of $A(p)$ even at the expense of increasing the order of $B(p)$.

Example 3.4 (The optimality of p in Theorem 0.4). Following Corollary 0.5, we do a random selector on $\Omega = \{1\}$, i.e. $x_i = 1$ and $\mathbb{E}\delta_i = \lambda = k/m$ then we shall assume that

$$\left(\mathbb{E} \left| \frac{1}{k} \sum_{i=1}^m \delta_i - 1 \right|^p \right)^{1/p} \leq C\sqrt{\frac{p}{k}} + \frac{f(p)}{k}$$

for some function $f(p)$. Here we choose $m = p$ and $k = ap$ for some very small a . Then we find that for every $1 \leq j \leq m$

$$\left| \frac{j}{k} - 1 \right| \binom{m}{j}^{\frac{1}{m}} \lambda^{j/m} (1-\lambda)^{1-j/m} \leq C\sqrt{\frac{m}{k}} + \frac{f(m)}{k}.$$

Let us first fix $j = \lceil \gamma m \rceil$ and assume that $\gamma \geq 1/4$ and $1/2^m < a \leq 1/8$. This gives $\frac{j}{k} \geq \frac{\gamma}{a} \geq \frac{1}{4a} \geq 2$ and hence

$$\left| \frac{j}{k} - 1 \right| \geq \frac{1}{8a}.$$

Note that $1 \leq \binom{m}{j}^{1/m} \leq 2$ so that we can not expect any help here. Thus we find

$$\frac{1}{16} a^{\gamma-1} (1-a)^{1-\gamma} \leq \frac{1}{8} a^{\gamma-1+1/m} (1-a)^{1-\gamma} \leq C a^{-1/2} + \frac{f(p)}{ap}.$$

Let us now fix $\gamma = 1/4$ and choose a such that

$$2Ca^{-1/2} \leq \frac{1}{16} \left(\frac{1-a}{a} \right)^{3/4}$$

or equivalently

$$32Ca^{1/4} \leq (1-a)^{3/4}.$$

However $a \leq 1/8$ implies $1-a \geq 7/8$. Thus

$$a \leq \left(\frac{7}{8} \right)^3 \frac{1}{(32C)^4}$$

will do. Then we find

$$\left(a^{1/4} \frac{(7/8)^{3/4}}{32}\right) p \leq f(p).$$

Choose $a = (7/8)^3/(32C)^4$. Then we have

$$(3.6) \quad \frac{c_0}{C} p \leq f(p)$$

for an absolute constant $c_0 = (7/8)^{3/2}/32^2$. This shows that one can not reduce the order of $f(p)$, as long as we keep $A(p) \leq C\sqrt{p}$ in (3.5).

Remark 3.5. In fact, Example 3.4 provides more information. Instead of fixing γ , by sending $\gamma \rightarrow 0$ and choosing $a \leq \gamma/2$ appropriately we can find a different behavior. Indeed, then we have $|j/k - 1| \geq \gamma/(2a)$ and

$$\frac{\gamma}{4} a^{\gamma-1} (1-a)^{1-\gamma} \leq C a^{-1/2} + \frac{f(p)}{ap}$$

and since $a < \gamma$ and $(1-\gamma)^{1-\gamma} \geq e^{-1}$ we need $8eC a^{-1/2} \leq \gamma a^{\gamma-1}$ or

$$a^{\frac{1}{2}-\gamma} \leq \frac{\gamma}{8eC}.$$

Note that $(\frac{\gamma}{8eC})^{2/(1-2\gamma)} \leq \gamma/2$ for $\gamma \leq 1$. Hence with

$$a \leq \left(\frac{\gamma}{8eC}\right)^{\frac{2}{1-2\gamma}}$$

we have

$$\frac{\gamma a^\gamma}{8e} p \leq f(p).$$

Put $a = (\frac{\gamma}{8eC})^{2/(1-2\gamma)}$. Then we obtain

$$\left(\frac{\gamma}{8eC}\right)^{\frac{1}{1-2\gamma}} Cp \leq f(p).$$

Optimizing the left hand side in γ , we obtain $2\gamma \log(8e^2C) - 2\gamma \log(\gamma) = 1$ and

$$\left(16eC \log \frac{8e^2C}{\gamma}\right)^{-1 - \frac{1}{\log \frac{8eC}{\gamma}}} Cp \leq f(p).$$

Since $\gamma \log \gamma \rightarrow 0$ as $\gamma \rightarrow 0$, we choose

$$\gamma = \frac{1}{2 \log(8e^2C)}.$$

In order to obtain a lower bound for $f(p)$, we need to assume $8C \geq 1$ so that $\gamma \leq 1/4$. This yields for $C \geq 1.5$,

$$(3.7) \quad f(p) \geq \frac{1}{32\sqrt{2}e^{3/2+2/e} \log(8e^2C)} p \geq \frac{p}{c_1 \log C}$$

for some absolute constant c_1 . Compare (3.7) with (3.6). The estimate (3.7) is better for large C . Let us now fix p and put $C = p^\alpha$. Example 3.3 shows that α has to be nonnegative. (3.7) implies that for $\alpha > 0$

$$f(p) \geq \frac{p}{c_1 \alpha \log p}.$$

In particular, for $C = \sqrt{p}/\log p$, we obtain $f(p) \geq 2c_1^{-1}p/\log p$, which recovers the best constants obtained in [17].

Example 3.3 and Remark 3.5 yield the following result.

Theorem 3.6. *Under the hypotheses of Theorem 0.4, assume that*

$$\left\| \sum_{i=1}^n x_i \right\|_p \leq A(p) \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^* + x_j^* x_j) \right)^{1/2} \right\|_p + B(p) \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p}$$

for some functions $A(p)$ and $B(p)$. Then we have

- i) *The best possible order of the lower bound for $A(p)$ is \sqrt{p} , which can not be improved even if the order of $B(p)$ is increased.*
- ii) *If $\Omega(p/\log p) = A(p) = O(p^\beta)$ where $\beta \geq 1$, then the best order of $B(p)$ is $p/\log p$.*

In the commutative case, i) was proved by Pinelis and Utev in [28]. To the best of our knowledge, ii) is new even in the commutative setting, which shows that the order $p/\log p$ of $B(p)$ can not be reduced at the expense of increasing $A(p)$ to any polynomial order. It was proved in [28] that if one reduces $B(p)$ to a constant then the best order of $A(p)$ is exponential in p and in this case $B(p)$ can not be improved even at the expense of increasing the order of $A(p)$. Unfortunately, it is still unclear what is the optimal choice of $B(p)$ for fixed $\Omega(\sqrt{p}) = A(p) = O(p/\log p)$. All we know is that $\Omega(p/\log p) = B(p) = O(p)$ and $p/\log p$ is sharp when $A(p) = \Theta(p/\log p)$.

4. APPLICATION TO COMPRESSED SENSING

In this section, we indicate how our Rosenthal type inequality applies to certain problems in Compressed Sensing Theory. Let us briefly recall the background here following [6, 8, 35]. We want to reconstruct an unknown signal $f \in \mathbb{C}^n$ from linear measurements $\Phi f \in \mathbb{C}^k$, where Φ is some known $k \times n$ matrix called the measurement matrix. The reconstruction problem is stated as

$$(4.1) \quad \min \|f^*\|_0 \quad \text{subject to} \quad \Phi f^* = \Phi f$$

where $\|f\|_0 = |\text{supp} f|$ is the number of nonzero element of f . Since this problem is computationally expensive, we consider its convex relaxation instead:

$$(4.2) \quad \min \|f^*\|_1 \quad \text{subject to} \quad \Phi f^* = \Phi f$$

where $\|f\|_p = (\sum_{j=1}^n |f_j|^p)^{1/p}$ denotes l_p norm throughout this section. Exact reconstruction means that the solutions to (4.1) and (4.2) are both equal to f . f is assumed to be s -sparse, i.e. $|\text{supp} f| \leq s$. We refer to [6, 35] for why (4.2) is a good substitute of (4.1). However, the restricted isometry property (RIP) on Φ is an extremely important tool for exact reconstruction due to Candes and Tao [7] (see also [5]). Let Φ_T denote the $k \times |T|$ matrix consisting of the columns of Φ indexed by T . The RIP constant Δ_s is defined to be the smallest positive number such that the inequality

$$C(1 - \Delta_s)\|x\|_2^2 \leq \|\Phi_T x\|_2^2 \leq C(1 + \Delta_s)\|x\|_2^2$$

holds for some number $C > 0$ and for all $x \in l_2$ and all subsets $T \subset \{1, \dots, n\}$ of size $|T| \leq s$. Candes and Tao proved the following theorem [5, 7]:

Theorem 4.1. *Let f be an s -sparse signal and Φ be a measurement matrix whose RIP constant satisfies*

$$\Delta_{3s} + 3\Delta_{4s} \leq 2.$$

Then f can be recovered exactly.

Since Δ_s is nondecreasing in s , in order to verify RIP, it suffices to show that

$$\Delta_{4s} \leq \frac{1}{2}$$

or simply $\Delta_s \leq \frac{1}{2}$ by adjusting constant if necessary. In this section, we apply Corollary 0.5 to study the problem of reconstruction from Fourier measurements. Two cases will be considered. In the first case we fix the support T of f . In the second case we allow it to vary. In the following, C will always denote the constant in Corollary 0.5 and \mathbb{C}^m will always denote the m -dimensional complex Euclidean space equipped with l_2 norm.

Example 4.2 (Fourier measurements). We consider the Discrete Fourier transform $\hat{f} = \Psi f$ where Ψ is a matrix with entries

$$\Psi_{\omega, t} = \frac{1}{\sqrt{n}} e^{-i2\pi\omega t/n}, \quad \omega, t \in \{0, \dots, n-1\}.$$

We want to reconstruct an s -sparse signal $f \in \mathbb{C}^n$ from linear measurements $\Phi f \in \mathbb{C}^\Omega$, where $\Omega \subset \{0, \dots, n-1\}$ is a uniformly random subset with average cardinality k and the measurement matrix Φ is a submatrix of Ψ consisting of random rows with indices in Ω . This is the Fourier measurement matrix considered in [6, 8, 35]. We can formulate

this random subset precisely using the Bernoulli model. Let $(\delta_i)_{i=0}^{n-1}$ be a sequence of independent selectors with $\mathbb{E}\delta_i = k/n$, for $i = 0, \dots, n-1$. Then

$$\Omega = \{j : \delta_j = 1\},$$

and $k = \mathbb{E}|\Omega|$.

Let y_i be the i -th row of Ψ and T the support of f . Write y_i^T for the restriction of y_i on the coordinate in the set T . For $x, y, z \in \mathbb{C}^n$, we define the tensor $x \otimes y$ as the rank-one linear operator given by $(x \otimes y)(z) = \langle x, z \rangle y$. Then

$$\Phi^* \Phi = \sum_{i \in \Omega} y_i^T \otimes y_i^T = \sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T.$$

Let $x_j = n y_j^T \otimes y_j^T$. Then

$$\frac{1}{n} \sum_{i=0}^{n-1} x_i = id_{\mathbb{C}^T} = I_T \quad \text{and} \quad \|x_j\| = n \|y_j^T \otimes y_j^T\| = n \|y_j^T\|_2^2 \leq s.$$

The next proposition follows easily from Corollary 0.5.

Proposition 4.3. *Assume that the average cardinality of a random set Ω is $k = \varepsilon^{-2}s$. Then for $t\varepsilon \leq C$,*

$$(4.3) \quad \mathbb{P} \left(\left\| \frac{n}{k} \sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T - id_{\mathbb{C}^T} \right\| \geq t\varepsilon \right) \leq s e^{-t^2/(2C^2e)}$$

where $\|\cdot\|$ is the operator norm.

Define

$$H := id_{\mathbb{C}^T} - \frac{n}{|\Omega|} \sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T.$$

Then $\Phi^* \Phi = \frac{|\Omega|}{n} (I_T - H)$. By the classical Bernstein inequality, $k/2 \leq |\Omega| \leq 3k/2$ with high probability, see [8, Lemma 6.6]. Therefore, by choosing $t\varepsilon < 1$, we find that the matrix $I_T - H$ is invertible with high probability. The precise meaning of “high probability” will become clear in a moment. This proposition is an analog of [6, Theorem 3.1] and [35, Theorem 3.3] with a *single* set T . We compare our results with previous results in the following remark. It is easy to show that $\mathbb{P}(k/2 \leq |\Omega| \leq 3k/2)$ given by Bernstein’s inequality dominates $1 - s e^{-t^2/(2C^2e)}$ for the value of k given below. Hence we only need to consider (4.3) for the probability of success.

Remark 4.4. i) For a single set T our result is more general than previous results on the invertibility of $\Phi^*\Phi$ obtained by Candes, Romberg, and Tao in the breakthrough paper [6]. In particular, if we put $t\varepsilon = 1/2$ and $\varepsilon^{-2} = 8C^2e(M \log n + \log s)$ for some $M > 0$, then we obtain $k = c_M s \log n$ for some constant c_M and $I_T - H$ is invertible with probability at least $1 - O(n^{-M})$. This gives [6, Theorem 3.1]. Together with [6, Lemma 2.3], or following verbatim the end of the proof of [33, Theorem 4.2 (section 7.3)], we recover the main results of [6].

ii) Allowing arbitrary choices of k and p we recover [33, Theorem 7.3], and we would like to thank H. Rauhut for bringing this to our attention. His proof requires considerably more technology. Both proofs are based on the optimal constant in the noncommutative Khintchine inequality (used in Rudelson's lemma) which was discovered independently by the first named author and Pisier (see [30] for more historic comments). We believe that our proof is more direct. Moreover, he established the exact reconstruction results based on his version of (4.3) cited above, which shows that an estimate like (4.3) is the key to the exact reconstruction problem.

We now investigate the case with multiple choices of T . First, it is clear that (4.3) remains valid for polynomially many sets T . In general, we have

$$(4.4) \quad \mathbb{P} \left(\sup_{|T| \leq s} \left\| \frac{n}{k} \sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T - id_{\mathbb{C}^T} \right\| \geq t\varepsilon \right) \leq |S| s e^{-\frac{t^2}{2C^2e}}$$

where $|S|$ denotes the number of set T with $|T| \leq s$. Note that

$$\Delta_s = \inf_{\alpha > 0} \sup_{|T| \leq s} \left\| \alpha \sum_{i \in \Omega} y_i^T \otimes y_i^T - id_{\mathbb{C}^T} \right\|.$$

It follows that

$$\mathbb{P}(\Delta_s \geq t\varepsilon) \leq \mathbb{P} \left(\sup_{|T| \leq s} \left\| \frac{n}{k} \sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T - id_{\mathbb{C}^T} \right\| \geq t\varepsilon \right).$$

Assume $s \leq n/2$. Since $|S| \leq s \binom{n}{s} + 1 \leq s(ne/s)^s$, if

$$(4.5) \quad 2 \log s + s \log \frac{ne}{s} < \frac{t^2}{2C^2e},$$

then with probability at least $1 - s^2(ne/s)^s e^{-t^2/(2C^2e)}$ we can recover *all* s -sparse signal f from its Fourier measurements Φf . From here we are able to obtain different bounds for k and the corresponding probabilities of success. As an illustration, we have the following result.

Proposition 4.5. *Assume $s \leq n/2$. Let $M > 0$ be a precision constant and n be a large integer such that*

$$2 \log s + s \log \frac{ne}{s} < (M + 1)s \log \frac{n}{s}.$$

Then a random subset Ω of average cardinality

$$(4.6) \quad k = 8C^2 e(M + 1)s^2 \log \frac{n}{s} = c_M s^2 \log \frac{n}{s}$$

satisfies RIP with probability at least $1 - s^2 e^s (n/s)^{-Ms}$.

Proof. Put $t\varepsilon = 1/2$ in (4.4). Since $k = s\varepsilon^{-2}$, we obtain $t^2 = 2e(M + 1)s \log(n/s)$. Thanks to the assumption on n , (4.5) is true. Then

$$\mathbb{P}\left(\Delta_s \geq \frac{1}{2}\right) \leq s^2 e^s \left(\frac{n}{s}\right)^{-Ms}.$$

We have proved the assertion. □

Remark 4.6. We can relax the bound for k a little to obtain polynomial probability of success. Indeed, the same argument as Proposition 4.5 yields that a random subset Ω of average cardinality

$$(4.7) \quad k = 8C^2 e(M + 1)s^2 \log n = c_M s^2 \log n$$

satisfies RIP with probability $1 - s^{2-s} e^s n^{-Ms}$.

The good aspect of Proposition 4.5 is that k is linear in $\log n$. Unfortunately, this is weaker than Rudelson and Vershynin's results in [35] $k = O(s \log n \log(s \log n) \log^2 s)$ for fixed probability $1 - \varepsilon$ of success, which was strengthened to super-polynomially probability of success by Rauhut following their ideas, see [33]. These results are obtained by using deep Banach spaces techniques. We added our results just for comparison. Of course, simple applications of Khintchine's inequality are not expected to replace either majorizing measure techniques or the iterative methods of [35] for the uniform estimates required for RIP. It seems known in the compressed sensing community that the tails bounds alone are not good enough. To conclude this section, we restate a conjecture on the best bound of k , see [35] (and [33] for further background).

Conjecture 4.7. *A random subset $\Omega \subset \{0, 1, \dots, n - 1\}$ of average cardinality $k = O(s \log n)$ satisfies RIP with high probability.*

Acknowledgements We would like to thank W. B. Johnson for bringing [28] to our attention. We also thank K. Lee for helpful discussions on compressed sensing. After our work was completed, we learned from S. Dirksen that he also essentially obtained (0.5) in

his Ph.D. Thesis [13] using a different method in the UIUC analysis seminar on November 3, 2011. We are also grateful to H. Rauhut for his comments on section 4 after he read our paper from arxiv.org, which have improved our statements.

REFERENCES

- [1] R. Ahlswede and A. Winter, *Strong converse for identification via quantum channels*, Information Theory, IEEE Transactions on **48** (2002mar), no. 3, 569–579. ↑4
- [2] H. Araki, *Golden-Thompson and Peierls-Bogolubov inequalities for a general von Neumann algebra*, Comm. Math. Phys. **34** (1973), 167–178. MR0341114 (49 #5864) ↑6
- [3] G. Bennett, *Probability inequalities for the sum of independent random variables*, Journal of the American Statistical Association **57** (1962), no. 297, 33–45. ↑1, 6
- [4] D. L. Burkholder, *Distribution function inequalities for martingales*, Ann. Probability **1** (1973), 19–42. MR0365692 (51 #1944) ↑1
- [5] E. Candes, M. Rudelson, T. Tao, and R. Vershynin, *Error correction via linear programming*, Foundations of computer science, 2005. focs 2005. 46th annual ieee symposium on, 2005oct., pp. 668–681. ↑23
- [6] E. J. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), no. 2, 489–509. MR2236170 (2007e:94020) ↑2, 22, 23, 24, 25
- [7] E. J. Candes and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215. MR2243152 (2007b:94313) ↑23
- [8] ———, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425. MR2300700 (2008c:94009) ↑22, 23, 24
- [9] P.-A. Cherix, M. Cowling, P. Jolissaint, P. Julg, and A. Valette, *Groups with the Haagerup property*, Progress in Mathematics, vol. 197, Birkhäuser Verlag, Basel, 2001. Gromov’s a-T-menability. MR1852148 (2002h:22007) ↑16
- [10] B. Collins and M. Junge, *What is a noncommutative Brownian motion?*, Preprint (2011). ↑12
- [11] J. B. Conway, *A course in functional analysis*, Second, Graduate Texts in Mathematics, vol. 96, Springer-Verlag, New York, 1990. MR1070713 (91e:46001) ↑6
- [12] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Second, Applications of Mathematics (New York), vol. 38, Springer-Verlag, New York, 1998. MR1619036 (99d:60030) ↑10, 11, 12
- [13] S. Dirksen, *Noncommutative and Vector-valued Rosenthal Inequalities*, Preprint, 2011. Thesis (Ph.D.)–Delft University of Technology. ↑27
- [14] T. Fack and H. Kosaki, *Generalized s -numbers of τ -measurable operators*, Pacific J. Math. **123** (1986), no. 2, 269–300. MR840845 (87h:46122) ↑3
- [15] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans. Inform. Theory **57** (2011march), no. 3, 1548–1566. ↑4, 6
- [16] P. Hitczenko, *Best constants in martingale version of Rosenthal’s inequality*, Ann. Probab. **18** (1990), no. 4, 1656–1668. MR1071816 (92a:60048) ↑1

- [17] W. B. Johnson, G. Schechtman, and J. Zinn, *Best constants in moment inequalities for linear combinations of independent and exchangeable random variables*, Ann. Probab. **13** (1985), no. 1, 234–253. MR770640 (86i:60054) ↑1, 2, 22
- [18] M. Junge, *Operator spaces and Araki-Woods factors: a quantum probabilistic approach*, IMRP Int. Math. Res. Pap. (2006), Art. ID 76978, 87. MR2268491 (2009k:46118) ↑12
- [19] M. Junge, *Doob's inequality for non-commutative martingales*, J. Reine Angew. Math. **549** (2002), 149–190. MR1916654 (2003k:46097) ↑5
- [20] M. Junge and Q. Xu, *Noncommutative Burkholder/Rosenthal inequalities*, Ann. Probab. **31** (2003), no. 2, 948–995. MR1964955 (2004f:46078) ↑5, 14, 15
- [21] ———, *On the best constants in some non-commutative martingale inequalities*, Bull. London Math. Soc. **37** (2005), no. 2, 243–253. MR2119024 (2005k:46170) ↑5
- [22] ———, *Noncommutative Burkholder/Rosenthal inequalities. II. Applications*, Israel J. Math. **167** (2008), 227–282. MR2448025 (2010c:46141) ↑2, 5, 13, 14, 16
- [23] S. Kwapien and W. A. Woyczyński, *Tangent sequences of random variables: basic inequalities and their applications*, Almost everywhere convergence (Columbus, OH, 1988), 1989, pp. 237–265. MR1035249 (91c:60020) ↑1
- [24] F. Lust-Piquard, *Inégalités de Khintchine dans C_p ($1 < p < \infty$)*, C. R. Acad. Sci. Paris Sér. I Math. **303** (1986), no. 7, 289–292. MR859804 (87j:47032) ↑16
- [25] F. Lust-Piquard and G. Pisier, *Noncommutative Khintchine and Paley inequalities*, Ark. Mat. **29** (1991), no. 2, 241–260. MR1150376 (94b:46011) ↑16
- [26] S. V. Nagaev and I. F. Pinelis, *Some inequalities for the distribution of sums of independent random variables*, Theory of Probability and its Applications **22** (1978), no. 2, 248–256. ↑2
- [27] D. Petz, *A survey of certain trace inequalities*, Functional analysis and operator theory (Warsaw, 1992), 1994, pp. 287–298. MR1285615 (95c:15038) ↑6
- [28] I. F. Pinelis and S. A. Utev, *Estimates of the moments of sums of independent random variables*, Theor. Probability Appl. **29** (1985), no. 3, 574–577. ↑19, 22, 27
- [29] G. Pisier, *Non-commutative vector valued L_p -spaces and completely p -summing maps*, Astérisque **247** (1998), vi+131. MR1648908 (2000a:46108) ↑5, 14
- [30] ———, *Introduction to operator space theory*, London Mathematical Society Lecture Note Series, vol. 294, Cambridge University Press, Cambridge, 2003. MR2006539 (2004k:46097) ↑25
- [31] G. Pisier and Q. Xu, *Non-commutative L^p -spaces*, Handbook of the geometry of Banach spaces, Vol. 2, 2003, pp. 1459–1517. MR1999201 (2004i:46095) ↑3
- [32] N. Randrianantoanina, *Conditioned square functions for noncommutative martingales*, Ann. Probab. **35** (2007), no. 3, 1039–1070. MR2319715 (2009d:46112) ↑5
- [33] H. Rauhut, *Compressive sensing and structured random matrices*, Theoretical foundations and numerical methods for sparse recovery, 2010, pp. 1–92. MR2731597 ↑25, 26
- [34] H. P. Rosenthal, *On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables*, Israel J. Math. **8** (1970), 273–303. MR0271721 (42 #6602) ↑1
- [35] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math. **61** (2008), no. 8, 1025–1045. MR2417886 (2009e:94034) ↑22, 23, 24, 26
- [36] M. B. Ruskai, *Inequalities for traces on von Neumann algebras*, Comm. Math. Phys. **26** (1972), 280–289. MR0312284 (47 #846) ↑6

- [37] N. M. Temme, *Special functions*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1996. An introduction to the classical functions of mathematical physics. MR1376370 (97e:33002) ↑11
- [38] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, ArXiv e-prints (April 2010), available at 1004.4389. ↑6
- [39] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. MR1217253 (94c:46133) ↑11

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, IL 61801

E-mail address: `junge@math.uiuc.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, IL 61801

E-mail address: `zeng8@illinois.edu`